

Masthead Logo

Veterinary Diagnostic and Production Animal
Medicine Publications

Veterinary Diagnostic and Production Animal
Medicine

2019

Invited review: Completeness of reporting of experiments: REFLECTing on a year of animal trials in the Journal of Dairy Science

Charlotte B. Winder
University of Guelph

Katheryn J. Churchill
University of Guelph

Jan M. Sargeant
University of Guelph

Stephen J. LeBlanc
University of Guelph

Annette M. O'Connor
Follow this and additional works at: https://lib.dr.iastate.edu/vdpam_pubs
Iowa State University, oconnor@iastate.edu

Part of the [Animal Experimentation and Research Commons](#), and the [Large or Food Animal and Equine Medicine Commons](#)
See next page for additional authors

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/vdpam_pubs/134. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

This Article is brought to you for free and open access by the Veterinary Diagnostic and Production Animal Medicine at Iowa State University Digital Repository. It has been accepted for inclusion in Veterinary Diagnostic and Production Animal Medicine Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Invited review: Completeness of reporting of experiments: REFLECTing on a year of animal trials in the Journal of Dairy Science

Abstract

Reproducibility is an essential element of the scientific process, and it requires clear and complete reporting of study design, conduct, and analysis. In the human and animal health literature, incomplete reporting is associated with biased effect estimates. Moreover, incomplete reporting precludes knowledge synthesis and undervalues the resources allocated to the primary research. The Reporting Guidelines for Randomized Controlled Trials for Livestock and Food Safety (REFLECT) statement, published in 2010, is a checklist developed by expert consensus to provide guidance on what study elements should be reported in any intervention trial (designed experiment) involving livestock. The Journal of Dairy Science (JDS) has recently endorsed the use of reporting guidelines. To assess the status of reporting of controlled experiments in JDS and to provide a baseline for future comparison, we evaluated the reporting of 18 items from the REFLECT statement checklist in a sample of 137 controlled trials published in JDS in 2017. Two reviewers independently screened titles and abstracts for relevance and then evaluated a sample of 120 papers reporting controlled trials (experimental studies involving at least one intervention and one comparison or control group), using yes or no questions. Although some items, such as treatment details and statistical analysis, were well reported, other areas, including sample size justification, allocation concealment, blinding, study flow, baseline data, and ancillary analyses, were often not reported or were incompletely described. This work highlights the need for authors and reviewers to take advantage of guidelines and checklists for reporting. Adherence to reporting guidelines can help improve the completeness of reporting of research, expedite and better inform the peer-review process, increase clarity for the reader, and allow for knowledge synthesis, such as meta-analysis, all of which serve to increase the value of the work conducted.

Keywords

reporting guidelines, trials, transparency, reproducibility

Disciplines

Animal Experimentation and Research | Large or Food Animal and Equine Medicine

Comments

This is a manuscript of an article published as Winder, Charlotte B., Katheryn J. Churchill, Jan M. Sargeant, Stephen J. LeBlanc, Annette M. O'Connor, and David L. Renaud. "Invited review: Completeness of reporting of experiments: REFLECTing on a year of animal trials in the Journal of Dairy Science." *Journal of Dairy Science* (2019). DOI: [10.3168/jds.2018-15797](https://doi.org/10.3168/jds.2018-15797). Posted with permission.

Authors

Charlotte B. Winder, Katheryn J. Churchill, Jan M. Sargeant, Stephen J. LeBlanc, Annette M. O'Connor, and David L. Renaud

Completeness of reporting of experiments: REFLECTing on a year of animal trials in the *Journal of Dairy Science*. Winder et al., page xxxx. Reproducibility in science requires comprehensive and accurate reporting of study design, conduct, and analysis. This observational study examined the prevalence of reporting of the 18 objective items in the Reporting Guidelines for Randomized Control Trials in Livestock and Food Safety (REFLECT) in trials published in the *Journal* in 2017. We found that while some items were well reported, there is room for improvement. Authors and reviewers should employ guidelines and checklists appropriate for their study type.

INVITED REVIEW: Reporting in animal trials

**Completeness of reporting of experiments: REFLECTing on a year of animal trials in the
*Journal of Dairy Science***

**Charlotte B. Winder^{*§}, Katheryn J. Churchill[†], Jan M. Sargeant^{*†}, Stephen J. LeBlanc^{*},
Annette M. O'Connor[‡], David L. Renaud^{*}**

^{*} Department of Population Medicine, University of Guelph, 50 Stone Road East, Guelph,
Ontario, N1G 2W1

[†] Centre for Public Health and Zoonoses, University of Guelph, 50 Stone Road East, Guelph,
Ontario, N1G 2W1, Canada

[‡] Department of Veterinary Diagnostic and Production Animal Medicine, College of Veterinary
Medicine, Iowa State University, Ames, IA, 50011-3619, USA

[§] Corresponding author: Charlotte B. Winder, 50 Stone Rd E., Guelph, ON, Canada N1G 2W1

email: winderc@uoguelph.ca

ABSTRACT

Reproducibility is an essential element of the scientific process, and requires clear and complete reporting of study design, conduct, and analysis. In human and animal health literature, incomplete reporting is associated with biased effect estimates. Moreover, incomplete reporting precludes knowledge synthesis and undervalues the resources allocated to the primary research. The Reporting Guidelines for Randomized Controlled Trials for Livestock and Food Safety (REFLECT) statement, published in 2010, is a checklist developed by expert consensus to provide guidance on what study elements should be reported in any intervention trial (designed experiment) involving livestock. *The Journal of Dairy Science* has recently endorsed the use of reporting guidelines. To assess the status of reporting of controlled experiments in the *Journal* and to provide a baseline for future comparison, we evaluated the reporting of 18 items from the REFLECT statement checklist in a sample of 137 controlled trials published in the *Journal* in 2017. Two reviewers independently screened titles and abstracts for relevance, and then evaluated a sample of 120 papers reporting controlled trials (experimental studies involving at least one intervention and one comparison or control group), using yes or no questions. While some items, such as treatment details and statistical analysis, were well-reported, other areas including sample size justification, allocation concealment, blinding, study flow, baseline data, and ancillary analyses were often not reported or incompletely described. This work highlights the need for authors and reviewers to take advantage of guidelines and checklists for reporting. Adherence to reporting guidelines can help improve the completeness of reporting of research,

expedite and better inform the peer-review process, increase clarity for the reader, and allow for knowledge synthesis, such as meta-analysis, all of which serve to increase the value of the work conducted.

Keywords:

reporting guidelines,

trials,

transparency,

reproducibility

INTRODUCTION

Concerns about incomplete reporting of research studies have been noted for decades (Sandiford et al., 1961) and have been identified in trials in livestock (Burns and O'Connor, 2008; Sargeant et al., 2009a; Sargeant et al., 2009b; Haimmerl et al., 2012), food safety (Sargeant et al., 2009b), and veterinary medicine (Lund et al., 1998). Although a study may have been well designed and executed, if the details documenting this rigor are not made explicit in the publication, readers cannot be certain what was done, and it becomes difficult to reproduce the research. Studies which fail to report key design features, such as random allocation, exclusion criteria, or details of intervention protocols and outcome measures, are more likely to report a positive intervention effect (Wellman and O'Connor, 2007; Burns and O'Connor, 2008; Sargeant et al., 2009a; Sargeant et al., 2009b; Brace et al., 2010). Additionally, if details of the conduct of a study are not clearly reported, the experiment often cannot be included in later meta-analysis, even if the data are relevant and would otherwise contribute. Concerns around incomplete reporting prompted the development of evidence-based reporting guidelines, such as the

Reporting guidelines For randomized controlled trials for livestock and food safety (REFLECT), which was developed by a consensus of experts to improve the quality of reporting of trials in livestock species (Sargeant et al., 2010).

Research synthesis is a fundamental part of science. Replication and consistency of results among studies are important to be able to draw conclusions about the effects of interventions. Systematic reviews and meta-analyses are becoming a common approach to research synthesis in animal science. The purpose of systematic reviews is to synthesize information across multiple experiments or studies to yield a high level of evidence for a specific research question. Systematic reviews often also include a meta-analysis, which takes into account within- and among-study variation when calculating the summary effect of an intervention. Systematic reviews also assess the risk of bias based on study design elements to aid interpretation of the summary effect.

Incomplete reporting in primary studies is a common finding in systematic reviews in animal health research (Dzikamunkenga et al., 2014; Baltzell et al., 2015), and limits the ability to answer relevant questions (Winder et al., 2018; Ali Naqvi et al., 2018). A lack of detail in reporting results in studies with suitable intervention groups and outcomes being excluded from meta-analysis (Rodney et al., 2015; Haimmerl et al., 2017), thus reducing the utility and impact of these individual studies.

Reporting guidelines function as a minimum standard to ensure that details related to key items are included, which provides a framework to assist both authors and reviewers, and improves clarity for the reader. This may help the research reach a wider audience, and, if consulted early in the process, can facilitate study design by ensuring key design components are considered. It should be stressed that reporting guidelines are *not* meant to prescribe how

researchers should design or execute their study, nor are they intended to be used as an assessment of the methodologic quality of the trial; reporting guidelines are an assessment of comprehensive reporting. Experiments may have different types of treatments (e.g. nutrition, reproductive protocols, vaccination, etc.), which may influence the reader's interpretation of the importance of adherence to specific design elements. However, reporting of fundamental elements of the design and conduct of an experiment is always necessary in order to allow for assessment of the study's results. For example, blinding of outcome assessors may be more or less important, or practically achievable, depending on the treatment or outcome in question. Comprehensive reporting does *not* mean that authors must always use blinding, but rather, it means that authors should always report whether or not blinding was used.

Thus, reporting guidelines serve to inform the reader or reviewer of the details of how the trial was conducted in such a way that the work could *then* be assessed for bias. Studies may be well-reported but poorly conducted; these should be rejected at peer-review. Conversely, peer-reviewed publication does not necessarily ensure that all important elements are clearly reported in a paper. Without appropriate reporting, it is not possible for reviewers and readers to assess the potential for bias, and therefore interpretation and applicability of the results are limited. Conversely, comprehensive reporting helps to maximize the value of work undertaken, by providing end-users with sufficient information to fully understand the assess potential bias, and to synthesise results. Standardized reporting allows for better evaluation of research integrity and minimizes potential misinterpretation (White and Larson, 2015). An ethical as well as a financial argument can be made for ensuring that best use is made of resources such as research animals and public funding.

The use of reporting guidelines is now endorsed by the *Journal of Dairy Science* in their *Instructions for Authors* <https://www.journalofdairyscience.org/content/inst-auth> , with encouragement of submission of a completed checklist from the relevant guideline appended to the submitted manuscript. A collection of reporting guidelines for research in animals is available through the Menagerie of Reporting guidelines Involving Animals (MERIDIAN), <https://meridian.cvm.iastate.edu> , which includes REFLECT (O'Connor et al., 2010b; Sargeant et al., 2010) and Strengthening the Reporting of Observational Studies in Epidemiology – Veterinary Extension (STROBE-Vet) (O'Connor et al., 2016; Sargeant et al., 2016). Both these guidelines were developed using published methodological frameworks (REFLECT, O'Connor et al., 2010; STROBE-Vet, Sargeant et al., 2016), which were based on published strategies for developing health research reporting guidelines (Moher et al., 2010). In particular, the REFLECT guideline was developed by consensus of a group of experts in livestock species, considering research with animal health, production, or food-safety outcomes (O'Connor et al., 2010e). Both REFLECT and STROBE-Vet have been endorsed and published by several leading journals in food science and veterinary medicine (for REFLECT: *Journal of Food Protection* (O'Connor et al., 2010d; Sargeant et al., 2010b), *Journal of Swine Health and Production* (O'Connor et al., 2010e), *Journal of Veterinary Internal Medicine* (O'Connor et al., 2010c), *Preventative Veterinary Medicine* (O'Connor et al., 2010a), *Zoonoses and Public Health* (O'Connor et al., 2010b; Sargeant et al., 2010a), and is now endorsed by *Journal of Dairy Science*. In addition, guidelines specific to the conduct of systematic reviews (PRISMA, Moher et al., 2009), laboratory animal experiments (ARRIVE, Kilkenney et al., 2010), and diagnostic accuracy studies (STARD, Kostoulas et al., 2017) are also available. As well, suggested guidelines for particular subject areas have been published in the *Journal of Dairy Science* (e.g.

for studies with calves (Kertz and Chester-Jones, 2004) and in reproduction (Lean et al., 2016)), but these do not preclude the need to report general study characteristics for the relevant type of study. The MERIDIAN site also offers a Reporting Interface for Guidelines on Research (RIGOR) <https://aflex.vrac.iastate.edu/checklist/?t=Reflect> , where manuscripts may be uploaded to facilitate checklist completion and exported as a portable document format (PDF).

While there is some evidence that reporting has improved since the publication of reporting guidelines for human and animal trials, there is still much room for advancement, as many key features remain under-reported (Plint et al., 2006; Totton et al., 2018). Researchers may be unaware that reporting guidelines exist. For example, a survey of editors of veterinary journals showed only one-third had instructions for authors which referred to reporting guidelines (Grindlay et al., 2014). Employing reporting guidelines allows for a standardized examination of the methodologic soundness of submitted manuscripts, furthering the high standards needed to assure journal quality. The role of journals extending beyond being gatekeepers to driving methodologic change has been well argued (Erb, 2010; More, 2010), and it behooves progressive journals to be self-critical and to advance standards. In order to further the quality of reporting in the *Journal of Dairy Science*, it would be beneficial to understand the current status of reporting in the *Journal*, to identify areas of weakness and provide a benchmark for future examination. As controlled experiments (i.e. trials) constitute a large proportion of published work in the *Journal of Dairy Science*, we focused our assessment on these.

Objectives

The objective of this cross-sectional observational study was to describe completeness of reporting in a sample of animal trials published in the *Journal of Dairy Science* in 2017, using 18

objective items of the REFLECT statement checklist (O'Connor et al., 2010; Sargeant et al., 2010).

MATERIALS AND METHODS

As this work was a cross-sectional study, it is reported using the relevant STROBE-vet headings for observational studies (O'Connor et al., 2016; Sargeant et al., 2016). The original study protocol for this work is available through the University of Guelph Atrium (<https://atrium.lib.uoguelph.ca/xmlui/handle/10214/13095>).

Definitions

In this paper, we consider the term ‘trial’ to be synonymous with ‘experiment’ or ‘study’, in which the researcher controls the allocation of animals (or groups of animals) to an experimental treatment, also known as an ‘intervention’. Experimental interventions may include, but are not limited to, diets, dietary supplements, drug treatments, vaccinations, reproductive protocols, housing or management practices, or surgical interventions. These interventions, as well as analyses, may be at different levels, such as udder quarter, cow, pen, or farm.

Eligibility criteria

Publications in the calendar year of 2017 in *the Journal of Dairy Science* were searched, using criteria designed to maximize the likelihood of identifying relevant studies. Interventions (i.e. treatments) may be (but are not limited to) diet compositions or supplements,

pharmacological products, housing strategies, management practices, or surgical procedures. A relevant comparator group also had to be present, and may be an untreated control group, a placebo, or an alternative intervention.

Literature search and initial screening

The search was conducted on May 17, 2018 in MEDLINE via OVID (University of Guelph license) with the following string: (AB=(experiment OR study OR studies OR trial OR challenge) AND JN=(Journal of Dairy Science) AND limit to YR="2017"). Search results were exported into DistillerSR (Evidence Partners Inc., Ottawa, Ontario, Canada). All title and abstracts were independently screened by two reviewers (KJC and CBW) for relevance using the following criteria.

1. Does the title or abstract describe a primary study involving animals or groups of animals as the experimental unit? (No, reject; Yes, proceed; Unclear, acquire full text and re-screen)
2. Does the title or abstract describe an intervention study with at least one comparator group? (No, reject; Yes, include; Unclear, acquire full text and re-screen)

Only those papers reporting intervention studies (i.e. designed experiments) were considered relevant, i.e. review articles, observational studies, or previously published data were not relevant. The experimental unit had to be a live animal or groups of animals; studies on animal products or *in vitro* studies were not relevant.

Sampling method and reporting assessment

Two *post-hoc* modifications from the original protocol were made: to limit assessment to a sample of relevant studies, and to limit the assessment to the prevalence of reporting of REFLECT items and not to assess risk of bias using the Cochrane tool (Higgins et al., 2016). These change to limit the scope of the study were made to aid in clarity of communication of the results. To determine the number of papers to assess, we calculated sample size for estimation of a proportion (Dohoo et al. 2010) was based on an estimated prevalence of 40 % of included papers reporting a sample size justification (based on results of Totton et al., 2018) with an acceptable error (precision) of 10 %, resulting in a minimum of 105 papers.

Of papers included after initial screening, the first 20 were used to pre-test the reporting assessment form among all reviewers (KJC, CBW, DLR, JMS), where any incongruence in answers to screening questions were discussed to ensure thorough and complete understanding of the questions. An additional 100 papers were selected using a random number generator to select the first paper (1 to 3), after which every third paper was included until 100 papers were obtained, involving more than one pass of the un-included papers. After pre-testing in triplicate, reporting assessment was done independently by two authors (KJC, CBW, DLR, or JMS), with disagreements resolved by consensus. If reviewers were authors on included studies, an alternative reviewer was assigned.

Extracted descriptive characteristics of included trials consisted of: first author affiliation (department, institution, country), trial location (country), population, farm type, and study design (type of experiment: challenge or field trial; participant paths: cross-over design (including Latin Square) or parallel group). Parallel (between-subjects) trials have study units assigned to a single treatment, while in crossover (within-subjects) trials, animals act as their own controls and receive treatments in a specific order (Lund et al., 1994). Field trials are

defined as research conducted in a clinical or field setting (including research herds) which involve investigator control of study unit selection and treatment allocation, but with natural development of the disease or outcome, whereas challenge trials involve purposive exposure to a pathogen or surrogate (which may occur ahead of (therapeutic challenge), or after (preventative challenge), the intervention). For example, intramammary inoculation with *E. coli*, manipulation of ruminal pH to simulate an acidosis event, or administration of a lipopolysaccharide, constitute experimental challenges.

For papers reporting multiple experiments, a separate reporting assessment was done for each trial reported in the manuscript. The REFLECT checklist is included as supplementary Table 1. Reporting assessment was based on the survey questions designed by Totton et al. (2018), where items 1 and 3 to 19 (of 22) in REFLECT were rephrased into questions (Table 2). Items 2, 20, 21, and 22 (appropriate scientific background given; appropriate interpretation of results accounting for hypotheses, sources of bias, and multiplicity of analyses or outcomes; external validity; and general interpretation of results), were not included because these items involve judgement to determine whether items have been comprehensively discussed. Conversely, the included items only require a determination of whether information on the design element is present or absent. For example, no judgement is required to determine whether or not a study described the method of allocation sequence generation (item 8), in contrast to assessing whether the authors included an appropriate interpretation of the generalizability of the results (excluded item 21).

In addition to the subdivision of items 3, 5, 10, and 11 by Totton et al. (2018), we also subdivided items 6, 7, and 8. Items 3, 5, 10, and 11 were split by the previous authors as they concerned more than one piece of information, while items 6 and 7 were subdivided as it was

expected that few studies would have reported sample size justification, and item 8 was modified to include a question on random allocation reported elsewhere in the manuscript, and to capture those reporting a method of study unit allocation which was not random (e.g. systematic assignment). Two questions pertained to cross-over studies' reporting of washout periods, which was not specifically included in REFLECT. The denominator for all items was the total number of included trials, with the exception of washout period questions only pertaining to cross-over trials, and subdivisions of item 8 only pertaining to studies reporting random allocation of study units.

Statistical analysis and presentation of results

After consensus was achieved for all included studies, results were exported from DistillerSR into STATA/SE 15.0 (StataCorp, College Station, TX) where descriptive statistics were tabulated for all fields. Study characteristics and the comprehensive reporting assessment are reported as tables.

RESULTS AND DISCUSSION

Because this study was limited to describing completeness of reporting of aspects of trial design, implementation, and analysis, we consciously limit our speculation on reasons as to why items were not reported. In addition, it is important to note that completeness of reporting does *not* assess the risk of bias in the trials examined, although completeness of reporting is a prerequisite for assessment of risk of bias by the reader or reviewer. Specific design elements such as randomization or blinding may have greater or lesser importance depending on the nature

of the experiment, but completely reporting what was done is fundamental to all scientific papers. Comprehensive reporting is a pre-requisite in order for the reader to be able to undertake a subject-specific interpretation of the specific design elements.

Study population and descriptive characteristics

The inclusion of studies is summarized in Figure 1. Of the 595 articles found in the literature search, 230 (39 %) were included after relevance screening, and every third article was included until 120 papers sampled for study characteristics and assessed for reporting criteria, which reported 137 unique trials. Eleven papers contained multiple trials. Included studies are available as supplemental file S1.

The study characteristics are outlined in Table 1. Dairy cattle were the population in the majority of the sampled studies, although studies involving dairy calves or heifers, dairy goats, and dairy sheep were also represented. Nearly one-quarter of studies did not report the trial setting; of those which did report, research or university farms were most common. First author affiliations included 29 department types (e.g. dairy science, animal biology), from 66 institutions (including government, academia, and private industry), and 24 countries from 5 continents. Field trials were most common (131/137), which may be completed in commercial or research herds, and involve investigator control of study unit selection and intervention allocation, but not exposure to disease or outcome. Challenge studies accounted for the remainder (6/137), where the investigator controls study unit selection and intervention allocation, and there is purposive exposure to disease as described above. One-third (42/137) of the experiments used a cross-over design while two-thirds (95/137) had parallel groups.

Reporting assessment

The prevalence of reporting of REFLECT items is outlined in Table 2. Further discussion on assessment of items and explanation of the item are outlined below; a brief explanation of the importance of each item is based on the REFLECT Explanation and Elaboration document (Sargeant et al., 2010) which provides a more detailed explanation and examples for each item.

Title/abstract (1). Random allocation of study units was reported in 87 of the 137 included trials; an additional 17 studies reported this information elsewhere in the text. It is recommended that this study design information is included in the title or abstract to facilitate identification of the study in literature searches, for example when conducting systematic reviews. For this reason, REFLECT also recommends the use of the terms ‘challenge trial (/study, /model)’, and ‘field trial’ or ‘clinical trial’ in the title or abstract. The proportion of studies that reported random allocation of study units was substantially greater than a sample of livestock trials examined in 2008, in which only 26 % reported this (Sargeant et al., 2009a). However, it is unclear how many JDS manuscripts were included in that sample. Totton et al. (2018) similarly found that this item showed significant improvement (OR=1.97, 95% CI=1.23-3.16) since the publication of REFLECT, where the post-2010 sample had 17 of 28 studies reporting this item.

It should be noted that while ‘study units’ may not be clearly or consistently, or always even correctly defined in the dairy science literature (Bello et al., 2016), this item only captures whether or not study units were reported, not whether the study unit was defined or analysed appropriately. REFLECT includes the synonyms of ‘unit of concern’ or ‘experimental unit’; for example, a study unit may be a mammary quarter, animal, pen, or farm (Sargeant et al, 2010).

Methods – participants (3). Understanding potentially important differences between the trial and the target population is important to assessing external validity; while settings were well reported (126/137), eligibility criteria were less commonly reported at both the animal (75/137) and farm (6/137) levels. Conducting a trial in a university or research herd affords a degree of control over some variables which may improve internal validity, with the trade-off that the results may be less easily generalized to commercial circumstances. In some cases, use of a research facility may be necessary to determine the outcome, for instance this requires rumen fistulation or sacrifice for post-mortem sampling. Reporting is most complete if it explicitly states why the trial was conducted in a given setting. As two-thirds of our sample trials were done in university herds, we speculate that authors may not have seen the need to specify why the trial was conducted in their own institutional herd. There is evidence that studies failing to report inclusion and exclusion criteria are more likely to report a positive treatment effect in both livestock (Sargeant et al., 2009a) and companion animal trials (Sargeant et al., 2010).

Methods – interventions (4). Both details of interventions and the level (individual animal vs. pen, group, or farm) at which they were administered were commonly reported in our sample (134/137). The interventions in studies in our sample was almost always given at the individual animal level; cluster (group) allocation was uncommon. Details on intervention administration is mentioned in the *Journal's* Instructions for Authors document, which may be why a greater degree of compliance with this item was seen. This was similar to a review of reporting of trials on bovine respiratory disease (Totton et al., 2018), in which 93 % of the studies published after 2010 reported this item. Lack of compliance with this item has also been shown to be associated with reporting positive treatment effects (Sargeant et al., 2009a); it is encouraging that this item appears to be well reported in our sample.

Methods - objectives, hypotheses (5). All intervention studies should clearly state an objective and the hypotheses to be tested; while 127/137 explicitly reported study objectives, only 97/137 stated a hypothesis in terms of the null hypothesis. Understanding if the trial is designed to test superiority, non-inferiority, or equivalence will allow the reader to better interpret the results, and gather whether appropriate sample size calculations and statistical analyses were done. Similarly, while objectives were reported in 96 % of field trials examining antimicrobial treatments for bovine respiratory disease, hypotheses were stated in only 32 % (Totton et al., 2018). The Instructions for Authors in the *Journal* mention the need for an explicitly stated hypothesis, which may contribute to generally good reporting.

Methods - outcomes (6). Reporting of the primary and secondary outcomes was rarely done (13/137), which correlates with the absence of information about sample size rationale in the included studies (item 7). In a similar study of pre-harvest food safety intervention trials, the primary outcome was only defined in 10 % of abstracts examined in a 2009 study (Snedeker et al., 2012), and none in a 2008 evaluation of full papers (Sargeant et al., 2009b). The primary outcome can be determined several ways. Most commonly it is the outcome on which the sample size calculation was based or it may be explicitly stated in the trial protocol, while secondary outcomes are ones for which no explicit sample size was calculated. If a sample size for more than one outcome was calculated, the outcome with the largest required sample size required is referred to as primary. This allows the reader to judge whether the trial was appropriately powered to detect a meaningful difference.

The level at which outcomes were measured, and details of the methods used to enhance the quality of measurements was generally reported (127/137). The latter may include standards of laboratory testing, duplicated sampling, and training of outcome assessors.

Methods – sample size (7). This fundamental aspect of study design was infrequently reported, with 22/137 reporting a sample size justification, of which 16 of the 22 included a sample size calculation as justification. It should be noted that nearly all trials included in our sample had multiple outcomes; it is important to inform the reader as to which outcomes were considered in the calculation, and if any non-independence was accounted for. Sample size calculations allow a consideration of the appropriate power and sampling error for experimental studies, and are therefore considered an important component of methodological rigour in trial design (Latif et al., 2016). It is inconsistent with the principles of use of animals for research to conduct an experimental study involving animals that is either under- or over-powered. When reported, sample size calculations also allow the reader to know which hypothesis was primary. Sample size should consider the nature of the outcome data (e.g. continuous, dichotomous, time-to-event), the variability between study units in the outcome, the expected magnitude of the effect of treatment, and the desired level of confidence in the result and power of the study (Dohoo, 2004).

However, some studies do not use formal sample size calculation; for example, in our sample, 6 papers reported that sample size was determined based on use of animals enrolled in a concurrent or previous trial. Regardless, it is still important to report how the sample size was determined for transparency and to aid the reader in understanding power and the potential role of chance. This finding also illustrates how comprehensive reporting assessment differs from bias assessment. These six studies did report the rationale for the sample size and that is what REFLECT recommends. Our findings were similar to that of Totton et al. (2018), where approximately one-third reported how the sample size was determined. Totton et al. (2018) did not show a change over time for this item, although earlier work (albeit in a different field, small

animal clinical trials) found no studies in their sample reporting how sample size was determined (Lund et al., 1998). An evaluation of trials examining the efficacy of prostaglandin F_{2α} for treatment of bovine endometritis also found that none included a sample size calculation (Haimperl et al., 2012).

The *Journal* requires that authors state explicitly that IACUC (or equivalent animal ethics or animal use) approval was obtained prior to commencement of the study. In almost all cases, ethical approval would require the identification of a primary outcome and hypothesis, around which the sample size was calculated. Therefore, we suspect that the lack of compliance for this item stems from a lack of reporting, as opposed to a lack of inclusion of this aspect of study design.

Methods – washout period (N/A). This item is not part of the REFLECT statement, but was considered by the authors to be an important reporting aspect of cross-over trials; knowledge of the washout period length and justification is necessary to determine if the population was equivalent before the start of each block. Washout periods were reported in 21 of 45 cross-over studies, but only one trial reported a justification for washout period length.

Methods – randomization (8, 9, 10). Randomization is used to minimize baseline differences at the time of the intervention (de Boer et al., 2015). The REFLECT statement recommends that authors describe the method used to generate the random sequence, which presupposes that all studies are randomized. Random allocation of study units was reported in 104 studies, but only 7 of these reported the method used to generate the sequence. This is similar to a study of veterinary clinical trials (Lund et al., 1998), in which only 12 % reported on the method. This item about method of sequence generation is designed to ensure that authors are not incorrectly describing a study as random when in fact alternation methods or haphazard

methods of allocation were employed. Many of the assumptions for validity of statistical methods are reliant on the concept of exchangeable groups, which is established by randomization, and so it is critical that authors use the word random only when the approach to allocation actually is random. It is possible that random allocation was indeed used in more than 7 trials, but without explicit reporting this not possible to ascertain. Interestingly, this particular item was examined in a survey of controlled trials published in five veterinary journals, where an attempt to contact the authors of the trials was made to see if further details on randomization could be obtained; two-thirds of contacted authors were unable to provide further information (Di Girolamo et al., 2017). This highlights the need for complete reporting in the original publication.

Restrictions or blocking variables were reported in 51 of 104 studies, where variables were stratified and allocation occurred within the strata (e.g. sex, parity, weight category, etc.) to minimize differences in covariates between groups.

The method used to implement the random allocation sequence, specifying if it was concealed until interventions were assigned, was reported in 3 of 104 randomized trials. Only one trial reported who generated the allocation sequence and who enrolled study units. Bias may be introduced if the sequence of allocation is known by the person enrolling study units, especially if units have unequal value (for example, a farm manager enrolling dairy cattle might have conscious or unconscious preferences about their animals). This differs from blinding in that knowledge of the intervention group sequence (although blinded) may influence enrollment.

Methods – blinding (11). Whether or not blinding was employed was infrequently reported; 17/137 studies reported if blinding was implemented for any stage of the trial (those administering the intervention, caregivers, outcome assessors, or those analysing data). Blinding

is important for internal validity, because a lack of blinding has potential to influence, consciously or unconsciously, measures post-enrollment such as animal management and assessment of outcomes. If blinding is not possible at one or more level(s), this should be reported. It is important to state who is and is not blind to treatment groups. In some situations, blinding of outcome assessors may be less important, if the outcome can only be measured one way (such as mortality), or if the outcome is very objective. This could be interpreted differently than blinding of caregivers, where differential management of treatment groups may have substantial influence (Ribble, 1990). Our results differ from similar work examining reporting in field trials of antimicrobial therapy for bovine respiratory disease, where although blinding was infrequently reported for those administering interventions (25 %), caregivers (7 %), or data analysis (0 %), blinding was commonly reported for outcome assessors (86 %) (Totton et al., 2018).

Methods – statistical analysis (12). A clear explanation of the statistical analyses performed, including methods used to account for the organizational structure of the data, is necessary to assess internal validity of a trial. This item was almost universally reported (135/137), although it should be noted that this does not imply that all analyses were appropriate, only that the approach taken was reported and any accounting for non-independence was described. Some details on reporting of statistical analysis are included in the *Journal's* Instructions for Authors, which may be why a greater degree of compliance with reporting of this item was seen.

Results – study flow (13). The number of study units enrolled, receiving the intervention, completing the trial, and analysed should be reported. If deviation from the trial protocol or loss to follow up is relatively uncommon or simple (for short-term studies), this can be described in

the text, whereas a flow diagram is recommended to show the organizational structure for longer-term studies or those with complex design or multiple organizational levels (e.g. animals in pens in herds). The study flow was reported in half (68/137) of included studies, the majority of which were short-term interventions with data available in the text.

Results – recruitment (14). The dates when the trial took place were infrequently reported (51/137). Field conditions may be seasonal, and reporting when the trial occurred is important for the reader to place the trial population and results in an appropriate context.

Results – baseline data (15). Demographic and clinical characteristics of the trial population by intervention group were reported in 37/137 studies. Clinical characteristics are synonymous with biological measurements, such as rumen pH, somatic cell count, or haematological variables, etc., while demographic characteristic include, for example, breed, weight, sex, and parity. While randomization aims to generate comparable intervention groups, differences may occur due to chance, and the reviewer and reader should be presented with the group characteristics. Similar to our findings, only 37 % of studies evaluating pre-harvest food safety interventions reported baseline data (Sargeant et al., 2009b). Many studies did not report baseline characteristics by group, instead presenting an overall value and stating a non-significant test of baseline differences. Statistically testing for baseline differences is not recommended for randomized studies (de Boer et al., 2015), because truly random allocation means the differences must be due to chance. However, adjustment for baseline differences in analyses may be appropriate (Roberts and Torgerson, 1999). Failure to report baseline data for treatment groups has been associated with a greater proportion of positive treatment effects (Sargeant et al., 2010), highlighting the importance of reporting this item in order for the reader to appropriately interpret the study results.

Results – number analysed (16). The number of study units analysed in each group was reported for 78/137 studies. It is important to note that the recommendation is to report the study units in the denominator; it is up to the authors, reviewers, and readers to assess which unit is appropriate for the specific outcome (e.g. it may be the unit of allocation, the unit of observation, or the effective sample size). Stating such allows the reader to determine if the unit is appropriate, and to assess loss to follow up and protocol deviations. A similar prevalence (64 %) of reporting of this item was found in an assessment of trials examining antimicrobial therapies for bovine respiratory disease (Totton et al., 2018).

Results – outcomes and estimation (17). Effect sizes (i.e. the magnitude of effects of treatment) were commonly reported, but less so for summary level data, with 89/137 reporting both. In many cases it was not clear if the effect was unadjusted or adjusted for specified co-variables, but the precision of the estimate (e.g. standard error or 95 % confidence intervals) was generally well reported (131/137).

Results – ancillary analyses (18). Analyses were almost never (1/137) reported to be pre-specified or exploratory. Multiplicity of analyses increases the risk of a type I error, which may be additionally influenced by multiplicity of outcomes. Failure to report measurement of all outcomes described has been associated with a higher probability of reporting positive treatment effects (Sargeant et al., 2009a). It is not uncommon that additional (i.e., not pre-planned) analyses are done, perhaps based on initial results, but it is important to state this in order to appropriately interpret analyses as pre-planned or exploratory. For example, unplanned contrasts have a different interpretation than those pre-planned and for which the sample was derived. This said, determining if analyses were pre-specified generally requires *a priori* documentation, such as a time-stamped document published ahead of the experiment to state the planned experimental

protocol. Recommendations for protocols for human intervention trials, such as the Standard Protocol Items: Recommendations for Intervention Trials (SPIRIT) checklist have been published for guidance (Chan et al., 2012). In human medicine, trial protocols are increasingly required to be registered and publicly available (e.g. ClinicalTrials.gov) at the start of a study, and may be published in peer-reviewed journals. University repositories may be available to some research groups to publish time-stamped PDF documents, or we suggest that a time-stamped PDF protocol could be included as supplementary material when the experiment is published. This would allow for a clear delineation of outcome(s) and analyses were pre-planned to test a specific hypothesis, and prevent hypothesizing after the results are known, which increases the risk of a type I error, promotes inappropriate statistical practices, and potentially loses information about the original hypothesis (Kerr, 1998).

Results – adverse events (19). Few (10/137) studies reported adverse events. It is possible that this is because such events usually did not occur, however, without a statement as to whether such events occurred, it is not possible to be sure. While adverse events may not necessarily be the result of the intervention, they should be reported. If no adverse events were experienced, this should also be reported for transparency and because knowledge of adverse events is an important part of decision-making.

CONCLUSIONS

This work provides a benchmark of the completeness of reporting of key items for intervention trials published in the *Journal of Dairy Science* in 2017. Similar to other findings in both animal and human trials, while some items were well reported, there is room for

improvement in many areas. It is encouraging to see that several items appear to be very well-reported, such as details of experimental treatment administration and the approach to statistical analysis. It is interesting to note these items are clearly outlined in the *Journal's* Instructions to Authors, suggesting that further endorsement of reporting of design elements can result in improved reporting in the *Journal*.

We emphasize that adherence to reporting guidelines is not a measure of study quality, but serves to provide the reader or reviewer with the appropriate information in order to assess the work. Without transparent reporting, the reader or reviewer is unable to accurately assess areas of weakness or potential bias. The endorsement of reporting guidelines by the *Journal* should serve to encourage authors to consult such guidelines before embarking on a trial (to ensure appropriate considerations are taken in study design) and to submit the appropriate completed checklist with manuscript submission. Reviewers for the *Journal* should require authors to report fully and explicitly, which will expedite the review process and allow for identification of methodological flaws and appropriate interpretation of published studies. Adherence to reporting guidelines will serve to advance the *Journal's* high standards and increase the value of the research published within.

Transparency

Annette M. O'Connor and Jan M. Sargeant are co-leads and co-authors of the REFLECT and STROBE-Vet statement publications. Annette M. O'Connor hosts the MERIDIAN website and RIGOR interface through the Iowa State University. Stephen J. LeBlanc is a section editor of *Journal of Dairy Science*. The other authors declare no conflicts of interest.

REFERENCES

Ali Naqvi, S., D.B. Nobrega, P.E. Ronksley, H.W. Barkema. 2018. Effectiveness of precalving treatment on postcalving udder health in nulliparous dairy heifers: A systematic review and meta-analysis. *J. Dairy Sci.* 101:4707-4728.

Altman, D.G., I. Simera. 2016. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network. *J. Royal Soc. Med.* 109: 67-77.

Baker, M. 2016. 1,500 scientists lift the lid on reproducibility. *Nature* 533: 452-454.

Baltzell, P., T. Engelken, A.M. O'Connor. 2015. A critical review and meta-analysis of the magnitude of the effect of anthelmintic use on stocker calf production parameters in Northern US States. *Vet. Parasitol.* 214: 2-11.

Begg, C., M. Cho, S. Eastwood, R. Horton, D. Moher, I. Olkin, R. Pitkin, D. Rennie, K.F. Schulz, D. Simel, D.F. Stroup. 1996. Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *J. Amer. Med. Assn.* 276: 637-639.

Bello, N.M., and D.G. Renter. 2018. Reproducible research from noisy data: Revisiting key statistical principles for the animal sciences. *J. Dairy Sci.* 101: 5679-5701.

Bello, N.M., M. Kramer, R.J. Templeman, W.W. Stroup, N.R. St-Pierre, B.A. Craig, L.J. Young, E.E. Gbur. 2016. On recognizing the proper experimental unit in animal studies in the dairy sciences. *J. Dairy Sci.* 99: 8871-8879.

Brace, S., D. Taylor, A.M. O'Connor. 2010. The quality of reporting and publication status of vaccine trials presented at veterinary conferences from 1988 to 2003. *Vaccine.* 28: 5306-5314.

Burns, M. J., and A. M. O'Connor. 2008: Assessment of methodological quality and sources of variation in the magnitude of vaccine efficacy: A systematic review of studies from 1960 to 2005 reporting immunization with *Moraxella bovis* vaccines in young cattle. *Vaccine* 26: 144–152.

Chan, A.W., J.M. Tetzlaff, P.C. Gøtzsche, D.G. Altman, H. Mann, J.A. Berlin, K. Dickersin, A. Hróbjartsson, K.F. Schulz, W.R. Parulekar, K. Krleža-Jeric, A. Laupacis, D. Moher. 2012. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ* 2013;346:e7586.

de Boer, M.R., W.E. Waterlander, L.D.J. Kuijper, I.H.M. Steenhuis, J.W.R. Twisk. 2015. Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. *Int. J. Behav. Nutr. Phys. Act.* 12:4

Di Girolamo, N., M.A. Giuffrida, A.L. Winter, R. Meursinge Reynders. 2017. Reporting and communication of randomisation procedures is suboptimal in veterinary trials. *Vet. Rec.* doi:10.1136/vr.104035.

- Dohoo, I.R. 2004. The design of randomized controlled trials of veterinary vaccines. *Anim. Health Research Reviews*. 5:235-238.
- Dohoo, I., W. Martin, H. Stryhn. 2010. *Veterinary Epidemiologic Research*. VER Inc., Charlottetown, Canada.
- Dzikamunhenga, R.S., R. Anthony, J. Coetzee, S. Gould, A. Johnson, L. Karriker, J. McKean, S.T. Millman, S.R. Niekamp, A.M. O'Connor. 2014. Pain management in the neonatal piglet during routine management procedures. Part 1: A systematic review of randomized and non-randomized intervention studies. *Animal Health Res. Rev.* 15: 14-38.
- Erb, H.N. 2010. Changing expectations: Do journals drive methodologic change? Should they? *Prev. Vet. Med.* 97: 165-174.
- Grindlay, D.J. R.S Dean, M.M Christopher, M.L Brennan. 2014. A survey of awareness, knowledge, policies and views of veterinary journal Editors-in-Chief on reporting guidelines for publication of research. *BMC Vet. Research*. 10: 10.
- Haimperl, P., S. Arlt, W. Heuwieser. 2012. Evidence-based medicine: quality and comparability of clinical trials investigating the efficacy of prostaglandin F_{2α} for the treatment of bovine endometritis. *J. Dairy Research*. 79:287-296.
- Haimperl, P., S. Arlt. S. Borchardt, W. Heuwieser. 2017. Antibiotic treatment of metritis in dairy cows – A meta-analysis. *J. Dairy Sci.* 100:3783-3795.
- Higgins, J.P.T., D.G. Altman, P.C. Gøtzsche, P. Jüni, D. Moher, A.D. Oxman, J. Savović, K.F. Schulz, L. Weeks, J.A.C. Sterne. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. 343:d5928 doi: 10.1136/bmj.d5928
- Higgins J.P.T., J.A.C. Sterne, J. Savović, M.J. Page, A. Hróbjartsson, I. Boutron, B. Reeves, S. Eldridge. A revised tool for assessing risk of bias in randomized trials. *Cochrane Methods*. *Cochrane Database of Systematic Reviews* 2016: 10 (Suppl 1). [dx.doi.org/10.1002/14651858.CD201601](https://doi.org/10.1002/14651858.CD201601).
- Kane, R.L., J. Wang, J. Garrard. 2007. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *J. Clin. Epidemiol.* 60:241-249.
- Kerr, N.L. 1998. HARKing: Hypothesizing After the Results are Known. *Person. Soc. Psych. Rev.* 2:196-217.
- Kertz, A.F., H. Chester-Jones. 2004. Invited review: Guidelines for measuring and reporting calf and heifer experimental data. *J. Dairy Sci.* 87: 3577-3580.
- Kilkenny, C., W.J. Browne, I.C. Cuthill, M. Emerson, D.G. Altman. 2010. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* 8(6), e1000412.

Kostoulas, P., S.S. Nielsen, A.J. Branscum, W.O. Johnson, N. Dendukuri, N.K. Dhand, N. Toft, I.A. Gardner. 2017. STARD-BLCM: Standards for the Reporting of Diagnostic accuracy studies that use Bayesian Latent Class Models. *Prev. Vet. Med.* 138: 37-47.

Latif, L.A., J.E.D. Amadera, D. Pimentel, T. Pimentel, F. Fregni. 2016. Sample size calculation in physical medicine and rehabilitation: A systematic review of reporting, characteristics, and results in randomized controlled trials. *Arch. Phys. Med. Rehabil.* 92:306-315.

Lean, I.J., M.C. Lucy, J.P. McNamara, B.J. Bradford, E. Block, J.M. Thomson, J.M. Morton, P. Celi, A.R. Rabiee, J.E.P. Santos, W.W. Thatcher, S.J. LeBlanc. 2016. Invited review: Recommendations for reporting intervention studies on reproductive performance in dairy cattle: Improving design, analysis, and interpretation of research on reproduction. *J. Dairy Sci.* 99:1-17.

Lund, E.M., K.M. James, J.D. Neaton. 1994. Clinical trial design: Veterinary perspectives. *J. Vet. Internal Med.* 8:317-322.

Lund, E.M., K.M. James, J.D. Neaton. 1998. Veterinary randomized clinical trial reporting: A review of the small animal literature. *J. Vet. Intern. Med.* 12:57-60.

Moher, D., B. Pham, A. Jones, D.J. Cook, A.R. Jadad, M. Moher, P. Tugwell, T.P. Klassen. 1998. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 352: 609– 613.

Moher D., A. Librati, J. Tetzlaff, D.G. Altman. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6(7):e1000097.doi:10.1371/journal.pmed1000097

Moher, D., K.F. Schulz, D.G. Altman. 2001. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Lancet* 357: 1191-1194.

Moher, D., K.F. Schulz, I. Simera, D.G. Altman. 2010. Guidance for developers of health research reporting guidelines. *PLoS Med.* 72: e10000217. Doi:10.1371/journal.pmed.10000217.

More, S.J. 2010. Improving the quality of reporting in veterinary journals: How far do we need to go with reporting guidelines? *The Vet. J.* 184: 249-250.

O'Connor, A.M., J.M. Sargeant, I.A. Gardner, J.S. Dickson, M.E. Torrence, C.E. Dewey, I.R. Dohoo, R.B. Evans, J.T. Gray, M. Greiner, G. Keefe, S.L. Lefebvre, P.S. Morley, A. Ramirez, W. Sischo, D.R. Smith, K. Snedeker, J. Sofos, M.P. Ward, R. Wills. 2010a. The REFLECT statement: Methods and processes of creating Reporting Guidelines for Randomized Controlled Trials for livestock and food safety. *Prev. Vet. Med.* 93: 11-18.

O'Connor, A.M., J.M. Sargeant, I. Gardner, J. Dickson, M. Torrence, C.E. Dewey, I. Dohoo, R. Evans, J. Gray, M. Greiner, G. Keefe, S. Lefebvre, P. Morley, A. Ramirez, W. Sischo, D. Smith, K. Snedeker, J. Sofos, M. Ward, R. Wills. 2010b. The REFLECT statement: Methods and

processes of creating reporting guidelines for randomized controlled trials for livestock and food safety. *Zoonoses and Public Health*. 57: 95 – 104.

O'Connor, A.M., J.M. Sargeant, I. Gardner, J. Dickson, M. Torrence, C.E. Dewey, I. Dohoo, R. Evans, J. Gray, M. Greiner, G. Keefe, S. Lefebvre, P. Morley, A. Ramirez, W. Sischo, D. Smith, K. Snedeker, J. Sofos, M. Ward, R. Wills. 2010c. The REFLECT statement: Methods and processes of creating reporting guidelines for randomized controlled trials for livestock and food safety. *J Vet Int Med*. 24: 57-64.

O'Connor, A.M., J.M. Sargeant, I. Gardner, J. Dickson, M. Torrence, C.E. Dewey, I. Dohoo, R. Evans, J. Gray, M. Greiner, G. Keefe, S. Lefebvre, P. Morley, A. Ramirez, W. Sischo, D. Smith, K. Snedeker, J. Sofos, M. Ward, R. Wills. 2010d. The REFLECT statement: Methods and processes of creating reporting guidelines for randomized controlled trials for livestock and food safety. *J Food Protection*. 73: 132-139.

O'Connor, A.M., J.M. Sargeant, I. Gardner, J. Dickson, M. Torrence, C.E. Dewey, I. Dohoo, R. Evans, J. Gray, M. Greiner, G. Keefe, S. Lefebvre, P. Morley, A. Ramirez, W. Sischo, D. Smith, K. Snedeker, J. Sofos, M. Ward, R. Wills. 2010e. The REFLECT statement: Methods and processes of creating reporting guidelines for randomized controlled trials for livestock and food safety. *J Swine Health Prod*. 18: 18-26.

O'Connor, A.M., J. M. Sargeant, I. R. Dohoo, H. N. Erb, M. Cevallos, M. Egger, A. K. Ersbøll, S. W. Martin, L. R. Nielsen, D. L. Pearl, D. U. Pfeiffer, J. Sanchez, M. E. Torrence, H. Vigre, C. Waldner, M. P. Ward. 2016. Explanation and Elaboration Document for the STROBE-Vet Statement: Strengthening the Reporting of Observational Studies in Epidemiology – Veterinary Extension. *Zoonoses Public Health* 63: 62-698.

Plint, A.C., D. Moher, A. Morrison, K.F. Schulz, D.G. Altman, C. Hill. 2006. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med. J. Aust*. 185: 263-267.

Ribble, C.S. 1990. Assessing vaccine efficacy. *Can. Vet. J*. 31:679-681.

Roberts, C., and D.J. Torgerson. 1999. Baseline imbalance in randomised controlled trials. *BMJ*. 319: 185.

Rodney, R.M., P. Celi, W. Scott, K. Breinhild, I.J. Lean. 2015. Effects of dietary fat on fertility of dairy cattle: A meta-analysis and meta-regression. *J. Dairy Sci*. 98:5601-5620.

Sandifer, M.G., R.M. Dunham, K. Howard. 1961. The reporting and design of research on psychiatric drug treatment: a comparison of two years. *Psychopharmacol. Serv. Cent. Bull*. 1: 6-10.

Sargeant, J.M., R. Elgie, J. Valcour, J. Saint-Onge, A. Thompson, P. Marcynuk, K. Snedeker. 2009a. Methodological quality and completeness of reporting in clinical trials conducted in livestock species. *Prev. Vet. Med*. 91: 107–115.

Sargeant, J.M., J. Saint-Onge, J. Valcour, A. Thompson, R. Elgie, K. Snedeker, P. Marcynuk. 2009b. Quality of reporting in clinical trials of pre-harvest food safety interventions and associations with treatment effect. *Foodborne Path. Dis.* 6:989-999.

Sargeant, J.M., A. M. O'Connor, I. A. Gardner, J. S. Dickson, M. E. Torrence and consensus meeting participants, I. R. Dohoo, S. L. Lefebvre, P. S. Morley, A. Ramirez, K. Snedeker. 2010a. The REFLECT Statement: Reporting Guidelines for Randomized Controlled Trials in Livestock and Food Safety: Explanation and Elaboration. *Zoonoses Public Health* 57: 105-136.

Sargeant, J.M., A.M. O'Connor, I.A. Gardner, J.S. Dickson, M.E. Torrence, Consensus Meeting Participants: I.R. Dohoo, S.L. Lefebvre, P.S. Morley, A. Ramirez, K. Snedeker. 2010b. The REFLECT Statement: Reporting guidelines for randomized controlled trials in livestock and food safety: Explanation and elaboration. *J. Food Protection.* 73: 579 – 603.

Sargeant, J.M., A.M. O'Connor, D.G. Renter, D.F. Kelton, K. Snedeker, L.V. Wisener, E.K. Leonard, A.D. Guthrie, M. Faires. 2011. Reporting of methodological features in observational studies of pre-harvest food safety. *Prev. Vet. Med.* 98:88-98.

Sargeant, J.M., A.M. O'Connor. 2014. Issues of reporting in observational studies in veterinary medicine. *Prev. Vet. Med.* 113: 323-330.

Sargeant, J.M., A.M. O'Connor, I.R. Dohoo, H.N. Erb, M. Cevallos, M. Egger, A. K. Ersbøll, S. W. Martin, L.R. Nielsen, D.L. Pearl, D.U. Pfeiffer, J. Sanchez, M.E. Torrence, H. Vigre, C. Waldner, M. P. Ward. 2016. Methods and Processes of Developing the Strengthening the Reporting of Observational Studies in Epidemiology – Veterinary (STOBE-Vet) Statement. *Zoonoses Public Health* 63: 651-661.

Schulz, K.F., I. Chalmers, R.J. Hayes, D.G. Altman. 1995. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *J. Amer. Med. Assn.* 273: 408-412.

Schulz, K.F., D.G. Altman, D. Moher. 2010. CONSORT 2010 Statement: Updated reporting guidelines for reporting parallel group randomized trials. *BMJ.* 340: 698-702.

Snedeker, K.G., P. Canning, S.C. Totton, J.M. Sargeant. 2012. Completeness of reporting in abstracts from clinical trials of pre-harvest interventions against foodborne pathogens. *Prev. Vet. Med.* 104:15-22.

Totton, S.C., J.N. Cullen, J.M. Sargeant, A.M. O'Connor. 2018. The reporting characteristics of bovine respiratory disease clinical intervention trials published prior to and following publication of the REFLECT statement. *Prev. Vet. Med.* 150: 117-125.

Wellman, N.G., and A.M. O'Connor. 2007. Meta-analysis of treatment of cattle with bovine respiratory disease with tulathromycin. *J. Vet. Pharmacol. Ther.* 30: 234-241.

White, B.J., and R.L. Larson. 2015. Systematic evaluation of scientific research for appropriateness of data analysis to improve clinical decision making. *JAVMA*. 2047:759-762.

Winder, C.B., C.L. Miltenburg, J.M. Sargeant, S.J. LeBlanc, D.B. Haley, K.D. Lissemore, M.A. Godkin, T.F. Duffield. 2018. Effects of local anesthesia or systemic analgesia on pain associated with cautery disbudding in dairy calves: A systematic review and meta-analysis. *J. Dairy Sci.* 101: 5411-5427.

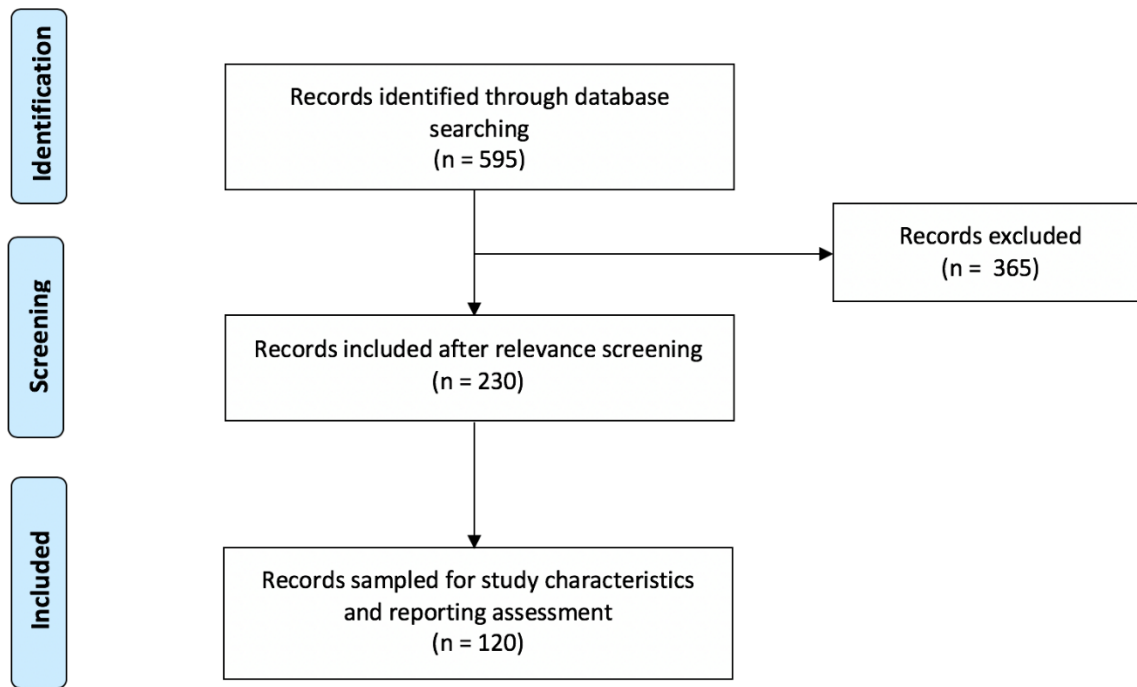


Figure 1. Flow of published articles through the initial database search, relevance screening, and sampling for extraction of study characteristics and reporting assessment of intervention trials published in *the Journal of Dairy Science* in 2017 for adherence to REFLECT items 1, and 3 to 19 (Sargeant et al., 2010).

Table 1. Characteristics of 137 intervention trials in animals from 120 selected papers published in 2017 in the *Journal of Dairy Science*.

Descriptive item		Proportion of studies in the sample
Study population ¹	Dairy cows	72 %
	Dairy calves/heifers	23 %
	Dairy goats	6 %
	Dairy sheep	1 %
Farm type	Commercial	13 %
	University/research	63 %
	Not reported	24 %
Study design: Experiment type	Experimental challenge ²	4 %
	Field trial ³	96 %
Study design: Participant paths	Cross-over	33 %
	Parallel	67 %

¹ Categories are not mutually exclusive

² Challenge studies involve investigator control of study unit selection, intervention allocation, and purposive exposure to disease or outcome

³ Field trials may be conducted in a clinical or field setting (including research herds) and involve investigator control of study unit selection and intervention allocation, but with natural exposure to disease or outcome

Table 2. Prevalence of reporting characteristics from 137 intervention trials in animals in 120 papers published in 2017 in *the Journal of Dairy Science*. Questions are taken from Totton et al. (2018), with additional subdivision of REFLECT¹ item 6, 7, and 8. An additional question was added to item 8 to capture studies reporting random allocation outside of title or abstract. An item specific to cross-over studies on reporting of washout periods was included (text in italics) but is not part of the REFLECT checklist.

REFLECT item number	Question	Prevalence (affirmative answer to the question; the information was reported)
1.	In the Title and/or Abstract, did the investigators report that the study units were randomly allocated to the interventions?	87 / 137 (64 %)
3.	In the Methods, did the investigators report eligibility criteria for the farm/owner/manager?	6 / 137 (4 %)
	In the Methods, did the investigators report study unit (animal or animal group) eligibility?	75 / 137 (55 %)
	In the Methods, was the setting where the data were collected described?	126 / 137 (92 %)
4.	In the Methods, did the investigators give precise details of the interventions intended for each group, the level at which the intervention was allocated, and how and when interventions were actually administered?	134 / 137 (98 %)
5.	Did the investigators report the specific objectives of the study?	127 / 137 (93 %)
	Did the investigators report the specific hypotheses of the study?	97 / 137 (71 %)
6.	Did the investigators give clearly defined primary and secondary outcome measures?	13 / 137 (9 %)
	Did the investigators report levels at outcomes were measured, and, when applicable, any methods used to enhance the quality of the measurements?	127 / 137 (93 %)
7.	Did the investigators provide a justification for sample size?	22 / 137 (16 %)
	Did the investigators provide a calculation for sample size?	16 / 137 (12 %)
	Were interim analyses reported?	1 / 137 (1 %)

	Were stopping rules used?	0 / 137 (0 %)
N/A	<i>Was a washout period reported?^{ba}</i>	21 / 45 (47 %)
	<i>Was an explanation for washout period length provided^b</i>	1 / 21 (5 %)
8.	Did the investigators report using a random allocation sequence at the relevant level of the organizational structure?	104 / 137 (76 %)
	Was the method used to generate the random allocation sequence reported?	7 / 104 (7 %)
	Were blocking factors or restrictions reported?	51 / 104 (49 %)
	Was allocation method reported but non-random (e.g. systematic)?	7 / 137 (5 %)
9.	Did the investigators report the method used to implement the random allocation sequence at the relevant level of the organizational structure, (e.g. numbered containers), clarifying whether the sequence was concealed until interventions were assigned?	3 / 104 (3 %)
10.	Did the investigators report who generated the allocation sequence?	1 / 137 (1 %)
	Did the investigators report who enrolled study units?	1 / 137 (1 %)
	Did the investigators report who assigned study units to their groups at the relevant level of the organizational structure?	1 / 137 (1 %)
11.	Did the investigators report whether those administering the interventions were blinded?	8 / 137 (6 %)
	Did the investigators report whether caregivers were blinded?	9 / 137 (7 %)
	Did the investigators report whether those assessing <u>all</u> outcomes were blinded?	11 / 137 (8 %)
	Did the investigators report whether those assessing <u>any</u> outcomes were blinded? ^c	17 / 137 (12 %)
	Did the investigators report whether those analyzing the data were blinded?	3 / 137 (2 %)
	Did the investigators report blinding (or the absence of blinding) at all?	17 / 137 (12 %)

12.	Were statistical methods used to compare groups for all outcome(s), and did the investigators clearly state the level of statistical analysis and methods used to account for the organizational structure (where applicable)?	135 / 137 (99 %)
13.	In the Results, did the investigators report the flow of study units through each stage for each level of the organization structure of the study?	68 / 137 (50 %)
14.	Did the investigators report dates defining the periods of recruitment and follow-up?	51 / 137 (37 %)
15.	Did the investigators report the baseline demographic and clinical characteristics of each group, explicitly providing information for each relevant level of the organizational structure?	37 / 137 (27 %)
16.	Did the investigators report the number of study units (denominator) in each group included in each analysis and the results in absolute numbers when feasible?	78 / 137 (57 %)
17.	Did the investigators report a summary of results for each intervention group, accounting for each relevant level of the organizational structure, and the estimated effect size and its precision (e.g. 95% confidence interval)?	89 / 137 (65 %)
18.	Did the investigators address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory?	1 / 137 (1 %)
19.	Did the investigators report all important adverse events or side effects in each intervention group?	10 / 137 (7 %)

^a Sargeant et al., 2010

^b Item not included in the REFLECT checklist

^c Includes blinding of any or all outcome assessors